

Report on the Data Quality Action Team's Activities

October 22, 2001

Executive Summary

In the six months since the formation of the Data Quality Action (DQA) Team, our team has worked with staff within the Information Technology Support Center (ITSC), with Headquarters and Area Statistical staff, and with other colleagues at Headquarters, Area Offices, and local facilities to accomplish the following:

- NPIRS data has been collected, organized, and inventoried.
- Critical elements of documentation of NPRIS data, database design, production processes, and programs has been produced and improved.
- Extensive, ongoing information about the DQA Team's activities has been shared with the entire I/T/U community on the IHS Internet site.
- Draft guidance has been fashioned as well as technical tools developed to allow authorized users easy access to NPIRS data.
- Final FY1998-2000 Workload and FY1999-2000 User Pop Reports have been produced.
- All 16 high priority and 7 additional medium to low priority data issues have been resolved and reflected in those reports.
- Current FY year-to-date Workload and User Pop Reports are being produced on a regular, monthly basis, providing Areas the opportunity to identify and resolve problems with their data well in advance of the final reports.
- A comprehensive reload of all NPIRS registration data is in progress and processing of these data will soon begin. This should result in markedly improved User Pop Reports.
- Work has begun to use Vality's Integrity software to improve patient unduplication methods for User Pop Reports.
- The design of a pilot implementation for a new, state-of-the-art national data repository to replace the current NPIRS database has been completed and data is being loaded and prepared for testing. This new design will allow for accurate statistical reports, markedly improved capabilities to provide Areas and local facilities information about the quality of their data, enhanced potential to perform GPRA, ORYX, and other clinical outcome measures, the creation of Area specific data marts, etc.
- Plans are being developed for future transitioning of these initiatives to regular IHS organizational units.

Background

At the end of last March, DIR and senior IHS management formed the Data Quality Action (DQA) Team, comprised of Headquarters East, Area Office, and Information Technology Support Center (ITSC) staff, including statisticians, an epidemiologist, system analysts, a physician/informaticist, and an Area Office management representative. They charged the team with conducting a full review of NPIRS processes, data movement, documentation, and procedures and then implementing changes to provide accurate, measurable, and timely improvements to national reporting systems.

To best address that charge, the team promptly established seven short-term goals:

- Complete an inventory of all NPIRS data
- Inventory and improve NPIRS documentation
- Coordinate with related ongoing data initiatives
- Implement an open, web-based system for communicating DQAT activities to all
- Provide access for appropriate users to the current NPIRS production database
- Implement a process for producing and iteratively improving Workload and User Pop Reports
- Take the first steps in implementing a true data warehouse/data mart design

During the intervening six months the DQA Team has worked with staff within ITSC, with Headquarters and Area Statistical staff, and with other colleagues at Headquarters, Area Offices, and local facilities to accomplish all seven of those goals. Although much could be said for each of the seven, this report focuses on progress made on the last three of these goals.

User Access to NPIRS

The DQA Team has worked with NPIRS staff to design and implement the technical pieces that will allow designated users to more easily access the information in NPIRS as well as restrict that access, where appropriate, to designated "views" or subsets of that information (e.g., restrict an Area Stat Officer to just his/her Area's data). Since the current designated "system owner" for the information in NPIRS is organizationally placed in the Office of Public Health (OPH), (Edna Paisano, Supervisory Statistician, Office of Program Statistics, OPH), we now await OPH's issuing the final guidance to give us permission and direction on how to specifically grant that access. We anticipate the guidance will specify to whom we can provide access; which categories of users are to be granted access to which subsets of data, which uses require individual review and pre-approval by the system owner and which fall within routinely approved uses that do not require individual request pre-approval, etc. In the meantime, NPIRS staff continues to provide access to these data, but only after individual review and approval by the current system owner of each request.

Improved Workload and User Pop Reports

The first week of October, NPIRS produced "final" FY1998-2000 Workload and FY1999-2000 User Pop Reports for Area review and approval. These reports included numerous major improvements in the data. During the late spring, summer, and early fall NPIRS staff worked closely with the Stat Officer Group and the DQA Team to identify, research, and implement "fixes" for the highest priority data problems affecting the accuracy of these reports. Of the

approximate 50 unique issues identified with Registration and APC/Inpatient workload data, 23 (16 high priority) have been resolved and reflected in the October reports. Of the remaining 27 open issues, the resolution of 4 are either being addressed by current processing efforts or are awaiting confirmation to close. Several members of the Statistical Officers Group and the DQA Team have already reviewed the resulting reports and many have noted marked improvements in their accuracy. Although problems remain that will, in turn and in order of priority, be addressed, there is no question that these reports are much, much better than those produced just six months ago.

In addition, in July, NPIRS began to regularly produce monthly, "year-to-date" versions of the Workload and User Pop Reports for the current fiscal year. These monthly reports will continue up to the production of the "final" reports in January of the following year. This production schedule affords Area Stat Officers the opportunity, if they choose, to review their data, and identify and address any problems well in advance of the production of the final reports. Also, this fall, the DQA Team plans to work with Cimarron Medical Informatics and NPIRS staff to harmonize, as much as possible, the PCC export and the NPIRS processing and workload report logic with the local PCC report logic so that the NPIRS and local PCC reports will be more comparable and the national reports easier for Area statistical officers to verify.

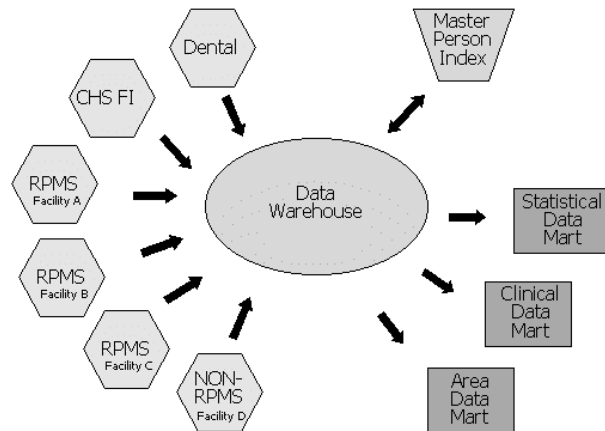
In early October, NPIRS began a comprehensive reload of its patient registration data from the field in order to improve the accuracy of its reports, especially the User Pop Reports. Although this project was primarily managed by NPIRS and other ITSC staff under the direction of the Patient Registration Workgroup, the DQA Team participated in its planning. Specifically, the DQA Team assisted ITSC programmers in the design and testing of the registration export and AIB processing programs, in identifying potential problems and suggesting solutions, in suggesting potential quality assurance checks to verify that data were good and the various processes involved in the reload processes were working appropriately, etc.

The DQA Team will be working with NPIRS to utilize Vality's Integrity software product, a complex software application that allows probabilistic matching, in the production of the FY 2001 User Pop Reports in January 2002. This product should allow us to more accurately unduplicate patients in the NPIRS database, resulting in more accurate User Pop Reports.

A New National Data Warehouse

While the DQA Team has been working with NPIRS to improve the accuracy and timeliness of current national reporting systems, it has also begun work with IBM to design and replace the current system with a new, state-of-the-art national repository. This national data warehouse (DW) will collect, store, and archive data almost exactly as it is received. These raw data can then be analyzed and useful information provided back to Areas and local facilities about their data. Subsets of data from this warehouse will, in turn, be extracted (and only then substantively transformed) to populate various data marts targeted to specific uses. Data marts that will be developed could and likely will include a statistical data mart from which Workload and User Pop Reports and other administrative reports can be derived; one or more clinical data marts from which ORYX, GPRA, Diabetes, or other outcome measurements can be derived; Area data marts from which Areas can obtain information that best meets their individual needs; and so forth.

National Data Warehouse



To test this concept, this summer we worked with IBM consultants to design the logical model and physical tables for a pilot data warehouse (PDW). At the same time we worked with ITSC programmers to design special export and export processing (AIB) programs. Registration and encounter data has already been gathered from 10 sites in 3 different Areas (9 RPMS and 1 non-RPMS site), from the Fiscal Intermediary (FI), from non-FI Contract Health Service systems, and from a separate dental information system. These data are now being extracted, minimally cleansed and transformed, and loaded into the new, pilot data warehouse for further analysis and testing.

This fall the DQA Team began work with the SAS Institute to design various reports to test the PDW, to demonstrate the utility of reports that can be produced from this design, and to begin to assess various specific strategies for improving the efficiency and effectiveness of the PDW design. For example, we are working with SAS to develop reports that would provide information to local sites about the timeliness of their data, less than expected counts based on historical norms, missing data in critical fields, erroneous or suspect codes, record counts, etc. We are also developing Diabetes and GPRA measure reports, exploring methods for unduplicating records about the same visits from different data sources, and exploring several different methods for setting workload flags and last visit dates to more easily allow true, "ad-hoc" statistical reports.

Later this fall and early winter we plan to work with Vality to use their probabilistic matching software to more accurately unduplicate patient registration records within and from different source databases (e.g., the same patient seen at different facilities) within this new data warehouse design.

Working with IBM, the DQA Team has produced a document that details our plans for moving ahead from the pilot phase to the full implementation of the first production phase of a national data warehouse. This user requirements/conceptual design document can be found on the IHS web page at the following address:

<http://www.ihs.gov/CIO/DataQuality/DW1Requirements1.pdf>

Future Plans and Directions

Upon completion of the official FY 2001 Workload and User Pop Reports in March 2002, it is anticipated that the normal production of these reports as well as the identification and resolution of the inevitable ongoing data issues will have been fully resumed by NPIRS. Also by then, the PDW implementation, including its evaluation, will have been completed. If the PDW is successful, ITSC will need to continue this initiative by progressively implementing a fully functional, system-wide national data warehouse (DW-1) that, together with its associated datamarts, will eventually replace the current NPIRS and ORYX databases. Finally, to accompany the work on improving data quality that is being done by ITSC, we hope that Areas and local facilities will also take progressively more significant and tangible steps within their own settings to address the data quality issues for which they are responsible and over which they have direct influence to, in their turn, help improve the quality of data in our many systems.